



# Applications des technologies sémantiques à la gestion des connaissances

## Sommaire

Sommaire .....	1
Introduction.....	2
1. Standardisation des termes et du vocabulaire d'un domaine .....	2
1.1 Vocabulaire GoodRelations pour le commerce électronique .....	4
1.2 Le vocabulaire SNOMED en médecine .....	5
1.3 Le vocabulaire web Dewey .....	6
2. Annotation sémantique et recherche de ressources.....	7
2.1 Annotation sémantique des ressources .....	7
2.2 Recherche sémantique des ressources.....	10
3. Processus de travail pour la gestion des connaissances .....	11
3.1 La modélisation des processus de travail .....	11
3.2 Un processus-cadre de gestion des connaissances.....	13
4. Web social et communautés de pratique.....	16
4.1 Le web social.....	16
4.2 Intégration du web social et du web sémantique .....	17
4.3 Ontologies pour le web social : FOAF, SIOC et folksonomies.....	19
4.4 Applications du web social sémantique .....	23
Conclusion.....	27

## Introduction

À la suite de l'introduction des principaux concepts de la modélisation, de la gestion des connaissances et des technologies sémantiques, nous allons maintenant présenter, sans entrer dans les détails techniques qui seront examinés dans les modules suivants, quatre applications génériques des technologies sémantiques à la gestion des connaissances, à l'aide d'exemples sélectionnés dans certains domaines de la connaissance. Nous qualifions ces champs d'application de génériques, car les mêmes concepts, processus ou techniques peuvent être appliqués dans d'autres domaines que ceux présentés ici.

Les technologies sémantiques peuvent servir à standardiser les termes et les concepts d'un domaine, facilitant ainsi la communication et le partage des connaissances. Ces technologies permettent aussi d'annoter sémantiquement un ensemble de ressources pour effectuer des recherches plus « intelligentes » et trouver des ressources répondant à certains besoins. Les technologies sémantiques sont également nécessaires pour soutenir un élément central de la gestion des connaissances : les processus de travail et de gestion au sein d'une organisation. Enfin, un quatrième aspect important de la gestion des connaissances est la mise en place de communautés de pratique gérant des processus informels de travail et d'apprentissage, ce qui implique un certain mariage des technologies du web social et du web sémantique.

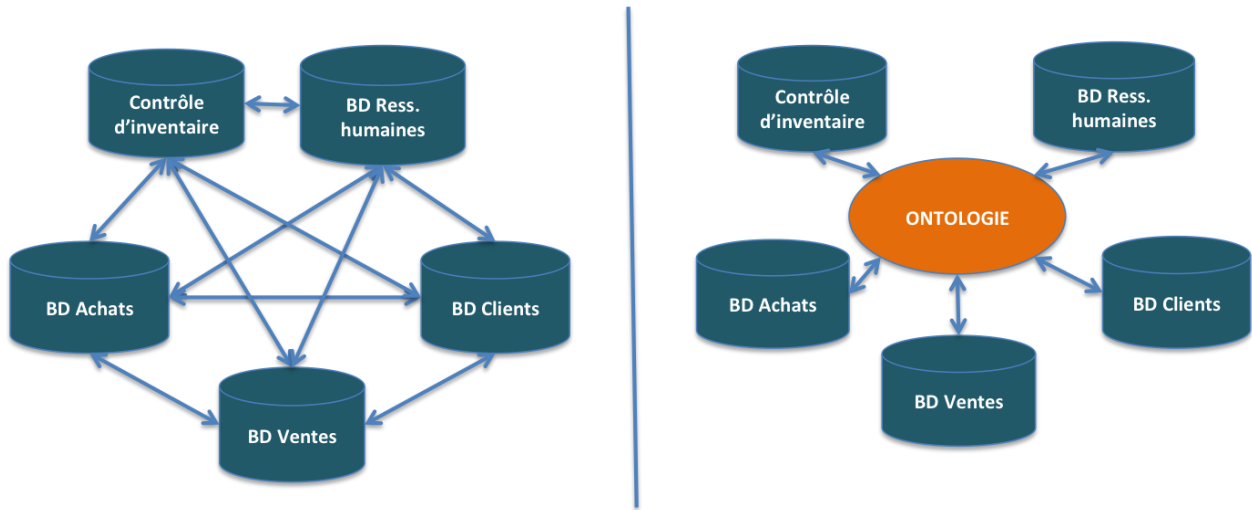
### 1. Standardisation des termes et du vocabulaire d'un domaine

Un problème important en gestion des connaissances consiste à partager l'information au sein d'une grande organisation ou entre plusieurs organisations. Ces organisations et leurs départements internes utilisent souvent différents vocabulaires, ce qui créera des difficultés d'interprétation dans la communication humaine. De plus, des terminologies différentes seront implantées dans les bases de données compliquant leur intégration informatique.

Au-delà du besoin de toute organisation d'intégrer ses propres systèmes d'information, de nombreuses motivations amènent l'organisation à partager ses connaissances avec d'autres organisations. Le cas le plus évident est la nécessité de coopérer avec d'autres dans la chaîne de production des produits et des services, de l'extraction des matières premières jusqu'au consommateur, ce qu'on appelle le e-business. Dans d'autres cas, il peut s'agir de partager des résultats de recherche ou de collaborer à des recherches précompétitives, dans la e-science. Parfois ces domaines interagissent comme dans le domaine médical où le même vocabulaire peut être utilisé dans un environnement clinique, dans le partage d'information avec les compagnies d'assurances ou avec les laboratoires de recherche.

Dans tous les cas, le problème est le même : le besoin d'un vocabulaire partagé par les personnes qui travaillent dans un même secteur industriel, professionnel ou gouvernemental. La plupart du temps, ces vocabulaires sont créés et maintenus par un organisme de standardisation, très souvent en langage naturel, en anglais, ou dans des versions multilingues. Ces définitions informelles peuvent impliquer des redondances, des inconsistances ou même des incompréhensions lorsque des personnes ou des groupes interprètent différemment les termes du langage utilisé.

Des problèmes de polysémie et de synonymie se présentent lorsque des bases de données utilisent un même terme ayant différentes significations ou différents termes ayant la même signification. L'importance des technologies sémantiques pour l'intégration d'information tient justement à la capacité de créer des ontologies qui « englobent » d'une certaine façon les différents schémas de bases de données et le sens donné aux termes qui y figurent.



**Figure 1** Utilisation d'un vocabulaire commun pour l'intégration d'information.

La figure 1 illustre cette idée. Le schéma de gauche montre cinq bases de données d'une organisation typique. Ces bases peuvent utiliser en partie des termes similaires. Leur usage conjoint exige de les interconnecter en examinant les correspondances appropriées entre les termes semblables pour résoudre les problèmes mentionnés plus haut. Il y a donc potentiellement neuf correspondances à examiner deux à deux dans le schéma de la figure 1. Inutile de dire que le problème illustré ici est multiplié lorsqu'il s'agit d'organisations différentes désirant lier leurs bases d'informations partiellement ou totalement.

Le fait de construire une ontologie intégratrice (figure de droite) qui utilise un vocabulaire de termes communs réduit à cinq le nombre de correspondances à examiner, les termes de chaque base de données étant associés à ceux de l'ontologie intégratrice, laquelle sera par ailleurs bien structurée avec une sémantique claire. Cette idée de l'utilisation d'un langage commun n'est pas nouvelle. Mais, ici, l'intégration se fait d'un point de vue sémantique, par l'examen de l'ensemble des concepts d'un domaine, ce qui pourra permettre par la suite d'y associer de nouvelles bases d'information.

Nous allons maintenant présenter deux exemples très différents où une ontologie est utilisée pour stabiliser le vocabulaire d'un domaine, facilitant le partage d'information sur le web entre une variété d'organisations. Le premier concerne le domaine du e-commerce et le second le secteur médical.

## 1.1 Vocabulaire GoodRelations pour le commerce électronique

GoodRelations (GR) est un vocabulaire web pour le commerce électronique (e-commerce), publié sous la forme d'une ontologie<sup>1</sup>. Développé par Martin Hepp à l'Universitaet der Bundeswehr Muenchen à partir de 2008, il a été utilisé pour annoter les pages web qui décrivent un produit ou un service dans un grand nombre de domaines. Un an après sa publication, des dizaines de pages étaient annotées selon le vocabulaire GR à l'aide du format de sérialisation RDFa. Le vocabulaire est aussi disponible sous d'autres formats de sérialisation tels que RDF/XML, Turtle ou Json et le tout nouveau format de microdonnées HTML5.

Ces annotations ont été utilisées par les moteurs de Yahoo! et de Google pour optimiser les recherches. Plus le produit est décrit précisément, plus il « sortira » comme résultat de recherche. Sur BestBuy ou overstock.com une page obtient un rang plus élevé que d'autres dans la mesure où elle est décrite précisément par l'ontologie GR. De plus, ces pages fournissent beaucoup plus d'informations standardisées, ce qui augmente les chances qu'un acheteur considère le produit.

Le but de GoodRelations est de définir un modèle de données pour le commerce électronique qui est neutre par rapport aux secteurs commerciaux. Il peut être utilisé aussi bien pour les produits d'électronique, les automobiles, les propriétés, les services professionnels ou tout autre type de produit ou de service. Les descripteurs du vocabulaire demeurent valides aux différentes étapes de la chaîne de valeur des produits, des matières premières jusqu'aux services après-vente.

Les classes principales du vocabulaire GR sont au nombre de quatre :

- `gr:BusinessEntity` : pour désigner l'agent, la personne ou l'organisation qui offrent le produit ou le service;
- `gr:Offering` : pour désigner une action, soit l'offre de vente, la réparation, la location ou pour exprimer de l'intérêt pour cette offre;
- `gr:ProductOrService` : pour désigner le produit ou le service;
- `gr:Location` : pour trouver un endroit où l'offre est disponible, une adresse web, un magasin, une bibliothèque, un bureau, etc.

S'ajoutent à ces classes principales 24 autres classes qui permettent notamment de spécifier les types d'agent, les modes de livraison, les prix, les conditions de licence, les garanties, etc. Le vocabulaire GR contient en outre :

- 44 propriétés de données qui permettent de décrire les classes, notamment la taille ou la couleur d'un produit, la date de disponibilité, etc.;
- 50 propriétés d'objets qui établissent des liens entre les individus des classes tels les modes de livraison disponibles ou le manufacturier;

---

<sup>1</sup> Site web : <http://purl.org/goodrelations/>

- des descriptions d'individus (ou de ressources) qui contribuent souvent à l'offre de produits et de services, tels les principales cartes de crédit (Visa, Mastercard, American express, etc.) ou les services de livraison tels que FederalExpress ou UPS.

L'ontologie GoodRelations est disponible sous licence Creative Commons. Un wiki<sup>2</sup> fournit une aide pour utiliser le vocabulaire, notamment plusieurs exemples de code RDFa à intégrer dans les pages web d'un produit pour le décrire à l'aide de l'ontologie GR.

## 1.2 Le vocabulaire SNOMED en médecine

En médecine et en biologie, les vocabulaires spécialisés sont souvent très étendus et complexes. Il s'agit donc d'un domaine naturel d'application des technologies sémantiques. Plus spécifiquement, en médecine clinique, SNOMED CT<sup>3</sup> est une ontologie largement utilisée créée en 2002 par la fusion de deux vocabulaires : SNOMED CT (Systematized Nomenclature of Medicine – Clinical Terms) du College of American Pathologists et le National Health Service Clinical Terms de Grande-Bretagne. Ce vocabulaire est désormais maintenu par l'IHTSDO (International Health Terminology Standards Development Organization) qui regroupait, en janvier 2012, 18 pays.

SNOMED CT est un très grand vocabulaire multilingue, le plus complet au monde dans son domaine. Il comptait déjà en 2011 plus de 311 000 concepts. Par exemple, le concept 22298006 correspond à l'infarctus myocardial. Tous les concepts de SNOMED CT sont structurés en taxonomies hiérarchiques acycliques. Par exemple, on y retrouve notamment les triplets suivants : Pneumonie virale EST-UNE Pneumonie infectieuse EST-UNE Maladie pulmonaire. Chaque concept peut avoir des parents multiples dans une taxonomie, par exemple, la pneumonie infectieuse et aussi une maladie infectieuse.

Les concepts de SNOMED CT sont liés par environ 1 300 000 liens ou relations de type (concept 1, relation, concept 2), les relations représentant un petit nombre de propriétés *comme* a-pour-cause ou localisation.

SNOMED CT contribue à l'amélioration des soins des patients en soutenant le développement et une recherche sémantique dans les dossiers de santé qui regroupent de l'information clinique sur des patients. Cela permet un accès efficace à l'information requise pour la prise de décision, les rapports et l'analyse de l'état du patient, ainsi qu'une meilleure communication entre patients et soignants.

SNOMED n'a pas été conçu originellement comme une ontologie; il est représenté à l'aide d'une logique de description. Depuis le développement de l'Ontology Web Language (OWL), il est possible de traduire le vocabulaire SNOMED en OWL. Une des possibilités offertes par la logique de description est la post-coordination qui permet de réduire le nombre de définitions de concept, bien que ce nombre demeure considérable. La post-coordination est cette fonction qui permet à un clinicien, par exemple, de créer de nouveaux concepts à partir de concepts existants. La vérification automatique de la consistance du vocabulaire est nécessaire périodiquement pour s'assurer que l'ensemble demeure cohérent.

---

<sup>2</sup> Site web : <http://wiki.goodrelations-vocabulary.org/Quickstart>

<sup>3</sup> Site web : <http://www.ihtsdo.org/snomed-ct/>

### 1.3 Le vocabulaire web Dewey

Le système de classification décimale Dewey (CDD), conçu dans les années 1870 par Melvil Dewey, définit une structure dynamique permettant d'organiser les ressources documentaires des bibliothèques. Après 23 éditions, disponibles en versions imprimées et en ligne, la CDD est le système de classification des ressources documentaires de bibliothèques le plus utilisé au monde.

Récemment, la coopérative internationale OCLC qui gère le système Dewey a entrepris de publier en partie les concepts du système de classification sous la forme d'un vocabulaire du web de données liées. L'organisme souligne le choix cornélien qui s'est offert à elle :

« une modélisation ontologique stricte avec un grand "O", ou des ensembles de données publiées qui ne sont que légèrement améliorés du point de vue sémantique (avec, ça et là, une pincée d'ontologie avec un petit "o"), mais qui peuvent être appliqués dans de nombreux domaines possibles. »

L'OCLC a opté pour le second choix. La dernière version des « Sommaires Dewey », c'est-à-dire les 1110 meilleures classes de la CDD 22<sup>4</sup>, a été publiée selon les principes des données liées, chaque classe étant identifiée par une URI et les données présentées de façon à pouvoir être réutilisées. Elle utilise un vocabulaire SKOS pour créer un modèle représentatif qui exprimerait les meilleures données de la CDD : identificateurs indépendants de la langue, terminologie multilingue et relations sémantiques. Le vocabulaire CDD est publié sous une licence Creative Commons facilitant la réutilisation des données de la CDD à des fins non commerciales.

Prenons l'exemple de la cote Dewey 641 et supposons que l'on veuille savoir à quoi elle correspond. On peut désormais utiliser un navigateur web standard et utiliser l'URI suivante : <http://dewey.info/class/641/>. Cette URI sous forme d'identificateur représente la classe 641 de la CDD. Elle redirige automatiquement le navigateur vers la page HTML de toutes les versions disponibles de cette classe, dans toutes les langues disponibles (<http://dewey.info/class/641/about>). La partie /about indique que cette URL contient une description générale du concept abstrait (c'est-à-dire la classe Dewey 641), et non le concept lui-même « Nourriture et boisson ».

Le concept lui-même, en tant qu'entité ou idée abstraite, ne possède pas de représentation pouvant être envoyée sur le web. Le serveur web oriente donc l'agent utilisateur vers un emplacement du web où cette entité est décrite. La possibilité de contourner la négociation de contenu en spécifiant directement le format souhaité existe toujours : <http://dewey.info/class/641/about.fr.rdf>.

Le vocabulaire web Dewey offre plusieurs avantages dont la facilité de passer d'une langue à l'autre ou de préciser la date de la version. L'affichage HTML est déjà enrichi du point de vue sémantique grâce à la norme RDFa du W3C. L'utilisation d'un navigateur qui prend le RDFa en charge (ou d'un extracteur RDFa) ouvre la voie vers de nouvelles possibilités de moissonnage, de collecte et de connexion des données Dewey. Ensuite, dewey.info présente

---

<sup>4</sup> Site web : <http://www.oclc.org/en-CA/dewey/resources/summaries.html>

une API simple qui utilise SPARQL, la technologie de recherche standard pour le web sémantique.

## 2. Annotation sémantique et recherche de ressources

En gestion des connaissances, la firme Gartner rapporte une tendance stable selon laquelle plus de 95 % du transfert d'information des humains vers l'ordinateur implique des textes en langue naturelle. D'un côté, les modèles de connaissances sont de plus en plus riches sémantiquement alors que de l'autre, les informations se trouvent dans des textes non structurés organisés hiérarchiquement.

*L'annotation sémantique* vise à résoudre ce problème en associant aux documents non structurés ou semi-structurés des références sémantiques, c'est-à-dire des URI « conceptuelles » présentes dans un modèle RDF ou une ontologie, soit des classes, des instances, des propriétés ou des relations. Le processus qui vise à ajouter à une ontologie de nouvelles instances (ou entités individuelles) extraites des textes est qualifié de *population de l'ontologie*.

### 2.1 Annotation sémantique des ressources

La tâche d'annotation sémantique est souvent subdivisée en deux étapes : l'examen de l'ontologie et la désambiguïsation.

L'examen permet d'identifier dans le graphe RDF ou dans l'ontologie un ou plusieurs concepts permettant de décrire des mots ou des fragments de texte d'un document. La désambiguïsation utilise de l'information contextuelle dans le texte ainsi que des connaissances dans l'ontologie pour choisir la bonne référence. Par exemple, si le texte parle d'un certain Éric Lapointe, dans une ontologie des personnes on trouvera possiblement deux instances, l'URI d'un sportif et une autre URI d'un chanteur. En consultant d'autres éléments du texte où on parle de football, on choisira le sportif.

Certains systèmes d'annotation sémantique effectuent aussi une population de l'ontologie en l'enrichissant de nouvelles instances membres de classes de l'ontologie. Par exemple, si un nouveau premier ministre est élu, un outil d'annotation des articles de journaux découvrira un nouveau nom de premier ministre et l'ajoutera comme instance de la classe des dirigeants d'États. Cette tâche est évidemment plus difficile qu'une simple annotation, encore plus si on demande à l'outil de découvrir de nouvelles classes ou relations.

L'annotation sémantique peut être réalisée manuellement, automatiquement ou encore semi-automatiquement. Dans ce dernier cas, les suggestions de l'outil d'annotation sont post-éditées par un expert humain. Puisque les annotations dépendent d'une ou de plusieurs ontologies, lorsque l'ontologie change, des annotations peuvent être perdues. Ce phénomène ainsi que la taille du contenu textuel sur le web fait en sorte que l'annotation manuelle est souvent impraticable, sauf dans des domaines et des applications limités. Elle permet toutefois de vérifier la qualité des méthodes automatiques et d'évaluer l'effort requis pour leur validation dans le cas de l'annotation semi-automatique.

L'extraction automatique d'information est nécessaire pour combler l'écart entre les textes non structurés et les graphes RDF ou les ontologies. L'extraction d'information à base d'ontologie se distingue de l'analyse du langage naturel classique en ce qu'elle utilise une ontologie formelle à la fois comme intrant et comme produit. Ce processus requiert la reconnaissance automatique des entités, des termes et des relations, à l'aide d'opérations courantes dans le traitement du langage naturel comme la lemmatisation.

Les outils manuels d'annotation sémantique doivent répondre à certaines conditions, outre le fait qu'ils doivent soutenir des références aux composantes d'une ontologie sous la forme d'URI. Comme l'annotation manuelle requiert beaucoup d'efforts, les outils devraient être collaboratifs pour permettre la distribution des efforts entre plusieurs annotateurs. Ils devraient aussi rendre possible l'annotation par les propriétés et leurs valeurs, et pas uniquement par les classes et les instances. Enfin, ces outils devraient pouvoir traiter plusieurs formats de document : HTML, PDF, XML, images et vidéos.

CREAM est un exemple d'un système élaboré pour l'annotation sémantique qui respecte les requis du paragraphe précédent. Il offre un éditeur de documents qui soutient la création d'annotations sémantiques pendant la création de documents. Il contient également un outil qui collige des entités RDF pertinentes déjà publiées sur le web de données et les rend disponibles pour l'annotation, réutilisant des instances existantes au lieu de les recréer. Ce système permet aussi d'intégrer des outils automatiques pour faciliter le processus d'annotation manuelle.

L'outil d'annotation manuelle dans CREAM s'appelle OntoMat. Il offre une interface utilisateur dans un navigateur web. Cet outil contient un navigateur de l'ontologie et des instances qui permettent aux utilisateurs de peupler l'ontologie avec de nouvelles instances. L'annotation de documents est faite en sélectionnant des fragments de texte, puis en appliquant un glisser-coller à la classe appropriée de l'ontologie, ou encore, une fois la classe choisie, en la portant sur une propriété ayant cette classe comme domaine, par exemple la date de naissance d'une personne.

L'exemple de la figure 2 montre une page web en cours d'annotation avec des classes telles que *Personne*, *Outil*, *Application*, *Événement*, *Projet*, *Publication*. La classe *Project* est sélectionnée et une de ses propriétés, *organizations\_involved*, est affichée. Une instance a été créée en sélectionnant d'abord le fragment de texte « University of Sheffield ». Le texte lui-même n'est pas sélectionné parce que CREAM utilise des triplets RDF qui sont indépendants du texte traité.



Figure 2 L'outil ONTOMAT d'annotation sémantique du système CREAM.

Un autre exemple d'un outil manuel d'annotation sémantique est intégré au système TELOS développé par le Centre LICEF de la TÉLUQ. TELOS contient un gestionnaire de ressources, un éditeur d'ontologies, un annotateur sémantique et un outil de recherche sémantique. Le gestionnaire de ressources permet d'importer des ontologies. L'éditeur d'ontologies (G-MOT) permet d'en créer ou d'en adapter de nouvelles. Ensuite, dans l'annotateur sémantique, appelé ici « référencier sémantique », on sélectionne une ou plusieurs ressources et une ontologie de domaine dans le gestionnaire de ressources.

On peut aussi choisir les ressources à référencer dans un scénario pédagogique, que ce soit un acteur, un document ou une activité. L'utilisateur peut voir les références sémantiques déjà associées à certaines ressources et les comparer. Ces références peuvent être des classes, des instances ou des propriétés de l'ontologie.

L'outil sert à ajouter, à modifier ou à supprimer des références aux ressources en les sélectionnant dans l'ontologie. Ces annotations sémantiques sont ensuite enregistrées de façon persistante dans la base de connaissances de l'ontologie technique qui gère les opérations de TELOS, ce qui revient à y ajouter des instances, à peupler l'ontologie. Dans le cas où on référence des acteurs, les connaissances qui lui sont associées constituent la base de son modèle utilisateur, lequel peut servir à lui offrir des conseils personnalisés.

## 2.2 Recherche sémantique des ressources

L'annotation sémantique des ressources permet aux utilisateurs de trouver tous les documents associés à des termes d'un graphe RDF ou d'une ontologie. La plupart des outils de recherche offrent également des fonctionnalités de navigation dans les ressources, ainsi que pour le raffinement des requêtes.

La recherche sémantique dans une banque de ressources est différente des moteurs de recherche classique comme celui de Google. Cela est dû à la structure de graphe RDF, ou de la structure plus élaborée d'une ontologie, reliant entre elles les annotations sémantiques et les fragments de texte auxquels ces annotations ont associées.

On peut distinguer divers types de moteurs de recherche sémantique. Swoogle<sup>5</sup>, par exemple, utilise des robots d'indexation ou collecteurs (*web crawlers*) pour rassembler à partir du web des documents RDF ou des documents HTML contenant du contenu RDF intégré, par exemple en format RDFa. Swoogle traite ces documents et leurs composants (termes et triplets) pour les enregistrer avec des métadonnées les décrivant dans sa base de données.

D'autres moteurs de recherche sémantique se concentrent sur les annotations et les utilisent pour trouver des documents, construisent des requêtes basées sur les ontologies ou offrent de naviguer dans les annotations selon la structure de l'ontologie. KIM<sup>6</sup> est une plateforme d'annotation sémantique pour la gestion des connaissances qui offre des fonctions pour la création de métadonnées, leur stockage et la recherche sémantique. L'annotation et la recherche sémantique se réalisent par la reconnaissance des entités en regard de l'ontologie KIM qui contient 250 classes et 100 propriétés à propos des personnes, des organisations, des produits, des lieux et des événements. La base de connaissance contient environ 200 000 entités.

Dans le système TELOS, le référencement sémantique des ressources a permis de créer des outils de recherche des ressources sur la base de leurs annotations sémantiques. Trois types de recherches sont disponibles.

- La *recherche simple* identifie les ressources par le nom des descripteurs sémantiques.
- La *recherche avancée* permet à l'utilisateur de créer des requêtes qui combinent des descripteurs des classes, des instances ou des propriétés de l'ontologie.
- La *recherche par ressource* affiche les ressources qui ont des référents sémantiques communs, identiques ou proches de ceux d'une ressource donnée. La recherche par proximité utilise la structure de l'ontologie utilisée dans les annotations.

Ce dernier type de recherche permet de trouver, par exemple, toutes les ressources dont les annotations sont proches de celles d'une activité en cours de conception ou d'apparier des acteurs selon leur profil de connaissances ou de compétences. En utilisant un concept de

---

<sup>5</sup> Site web : <http://swoogle.umbc.edu>

<sup>6</sup> Site web : <http://www.ontotext.com/kim>

compétence qui lie habileté générique et connaissances d'un domaine, les outils de recherche sont étendus à la recherche de ressources par compétence.

### 3. Processus de travail pour la gestion des connaissances

Le défi d'aider le personnel dans l'exécution de leurs processus de travail est au cœur de la gestion des connaissances dans les organisations. Ces processus sont, pour la plupart du temps, collaboratifs, donc multi-acteurs. La réalisation des processus de travail nécessite l'utilisation et le partage d'information pour atteindre un certain objectif, ainsi que la représentation du contexte dans lequel opèrent les utilisateurs.

Les outils de productivité utilisés dans les processus de travail contribuent à soutenir ces processus, mais offrent en général peu d'aide proactive. Par exemple, un outil comme Outlook fournit un agenda et une liste de contacts, mais laisse l'essentiel du travail à l'utilisateur. Par exemple, quand celui-ci inscrit « téléphoner à Jean Émond » dans une plage horaire de son agenda, le système ne fournit aucun lien automatique à la section des contacts pour que l'appel s'effectue en temps voulu. De plus, les informations contenues dans les outils de productivité sont régulièrement effacées. Par exemple, quand un employé reçoit un courriel de Jean Émond, avec des attachements décrivant une demande de service de la part d'un client, il peut l'effacer ou l'enregistrer dans un dossier, mais dans les deux cas, le lien entre les documents, le client ou Jean Émond est perdu.

Enfin, les systèmes actuels ne fournissent généralement pas d'information quant au contexte dans lequel chaque utilisateur travaille ou quant aux interactions qu'implique un processus de travail donné. Par exemple, si l'utilisateur est un avocat qui travaille sur six demandes de brevets différents, le système n'a aucune idée du brevet sur lequel il se concentre actuellement. De plus, le système ignore si l'utilisateur est en train de créer une demande, de réviser une demande préparée par le client ou de faire une recherche sur des brevets déjà enregistrés par d'autres entreprises. De telles informations permettraient au système informatisé d'aider l'utilisateur de façon proactive.

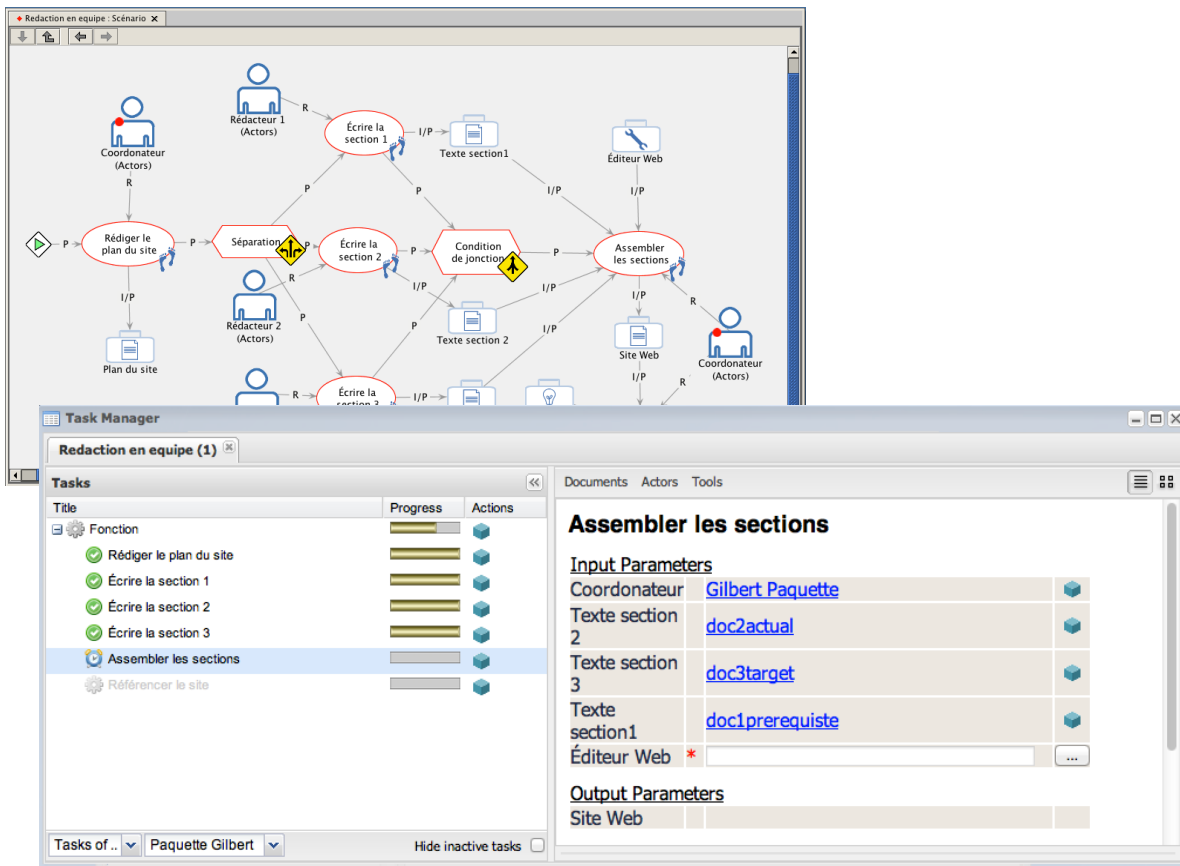
#### 3.1 La modélisation des processus de travail

Nous avons présenté plus tôt un exemple en langage MOT d'un processus de travail où quatre acteurs interagissent pour rédiger le contenu d'un site web sur un sujet donné. Lorsqu'il est exécuté dans un système comme TELOS<sup>7</sup>, capable de traiter les scénarios MOT, ou dans un moteur de *workflow* BPEL<sup>8</sup>, capable de traiter son équivalent en langage de notation BPMN, un tel modèle fournit aux utilisateurs désignés comme acteurs de ce processus, une vue du contexte, des interactions, des documents et des outils de travail.

---

<sup>7</sup> Paquette, G. (2010). An ontology-driven system for e-learning and knowledge management. Dans Paquette, G., [Visual Knowledge Modeling for Semantic Web Technologies: Models and Ontologies](#) Hershey, PA: IGI Global.

<sup>8</sup> Ouvans, C., Dumas, M., ter Hofstede, A. et van der Aalst, W. M. P. (2006). From BPMN process models to BPEL web services. Dans *Proceedings of Fourth International Conference on Web Services (ICWS 2006)*, Chicago, p. 85-292.doi:10.1109/ICWS.2006.67.



**Figure 3** Prise en compte du contexte dans un scénario de travail.

Outre le modèle graphique qui peut être affiché en permanence pendant l'exécution du processus, l'interface de la figure 3 montre les éléments du contexte dont le coordonnateur devra tenir compte au moment où il doit assembler les sections du site web. Il dispose du plan du site qu'il a rédigé au début du processus de travail et des documents des trois sections à assembler. Il lui reste à choisir un outil éditeur web avec lequel il créera le site web. Les menus de l'interface lui montrent les documents, les autres acteurs concernés et les outils utilisés. Le graphique et la liste des tâches (à gauche de la deuxième fenêtre) lui indiquent que sa tâche se situe après que les trois tâches d'écriture auront été accomplies. Les petits cubes donnent accès aux propriétés des tâches, des documents et des acteurs du processus, à leurs connaissances et compétences en jeu. Ces références sémantiques servent à fournir des conseils proactifs sur l'exécution des tâches, les interactions entre les acteurs ou le choix des documents à consulter.

De la même manière, on pourrait décrire le processus de travail d'une équipe chargée des brevets dans une grande organisation. On retrouverait dans le scénario un ordonnancement des tâches : le repérage des propositions de brevet auprès du personnel de l'entreprise (plusieurs cas possibles), la recherche dans les banques d'information pour des dépôts similaires (puis décider si on va de l'avant), la priorisation des demandes, la répartition de la rédaction des demandes, le suivi des demandes. On y retrouverait également plusieurs acteurs qui collaborent à ces tâches, un ensemble de documents sur la propriété

intellectuelle et les politiques de l'organisation, des outils de traitement des données, etc.

### 3.2 Un processus-cadre de gestion des connaissances

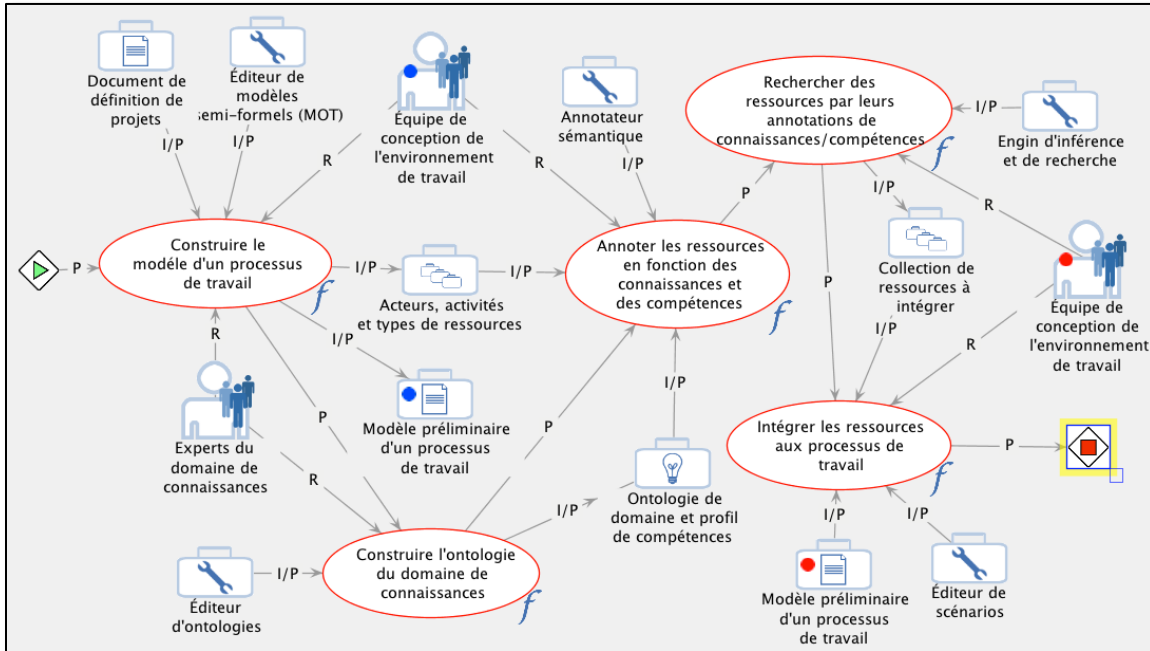
Nous présentons maintenant un processus-cadre d'implantation d'un projet gestion des connaissances dans une organisation. Celui-ci utilise trois des éditeurs G-MOT, l'éditeur de modèles MOT, l'éditeur d'ontologies et l'éditeur de scénarios.

Ce processus-cadre comporte trois grandes phases :

1. La rédaction d'un document de définition du projet de gestion des connaissances à partir d'un gabarit de cahier des charges. Cette phase consiste à appliquer certaines des méthodes de gestion des connaissances (section 4 du texte 1.2), notamment pour identifier les actifs intangibles et les connaissances critiques que l'organisation désire valoriser.
2. La construction d'un scénario de gestion des connaissances (GC) à partir du document de définition de projet. Cette phase implique le repérage et la modélisation des connaissances à valoriser ainsi que les processus de travail où implanter les activités de gestion des connaissances.
3. La génération d'un portail utilisateur donnant accès au scénario de GC, à ses activités et à ses ressources. Cette phase peut être réalisée automatiquement à partir du scénario de la phase 2 par un système comme TELOS ou le scénario peut être implanté à l'aide d'un outil de gestion de contenu comme WordPress, Joomla ou Drupal.

Nous allons nous concentrer sur la seconde phase. Celle-ci est décrite au moyen du processus de la figure 4. Ce scénario en langage MOT comporte cinq tâches, à la suite desquelles l'assemblage du portail prévu à la phase 3 pourra être réalisé.

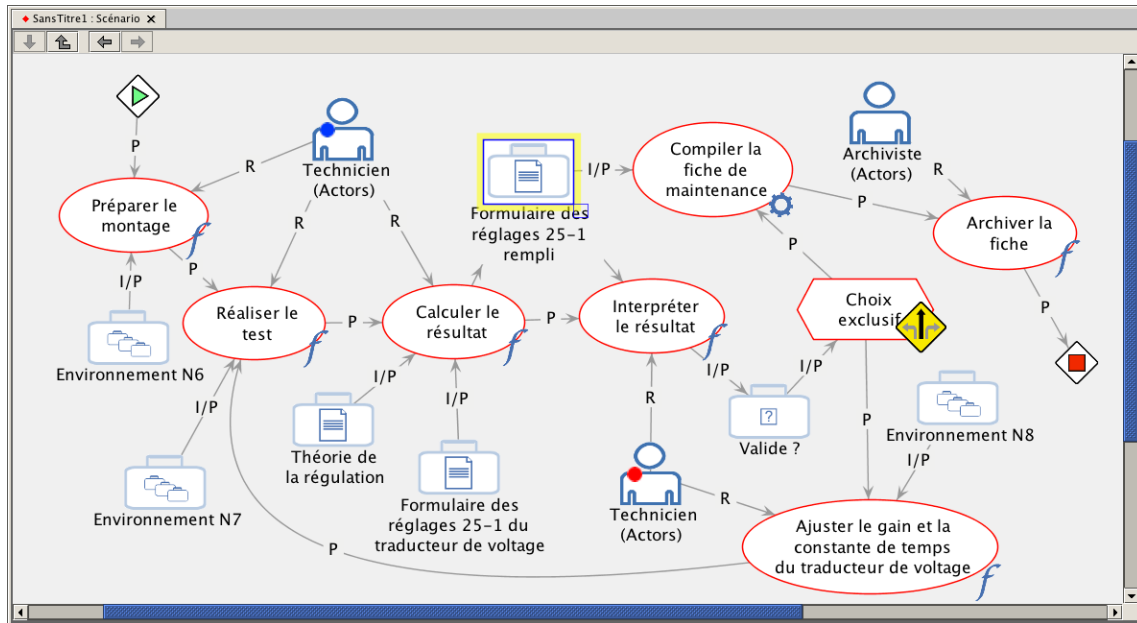
- La première tâche consiste à construire un *graphe semi-formel* (en langage MOT) *d'un processus de travail* spécifique dans lequel la gestion des connaissances sera intégrée. Cette tâche sera répétée plusieurs fois si plus d'un processus de travail est visé par le document de définition de projet. Ce graphe est réalisé à l'aide de l'éditeur G-MOT, à partir du document de définition du projet, par une collaboration entre une équipe d'experts du domaine de connaissances et l'équipe de conception de l'environnement de travail. Cette première tâche produit un (ou plusieurs) scénario semi-formel et la liste des ressources impliquées dans le scénario : acteurs, activités (tâches) et ressources (documents) à utiliser et à produire.
- La seconde tâche consiste, pour l'équipe d'experts du domaine, à *repérer et à modéliser les connaissances* (et au besoin les compétences) utilisées-crées par le scénario de travail. Grâce à l'examen des ressources du scénario semi-formel, l'équipe d'experts construit un modèle de connaissances sous forme d'ontologie ou de graphe RDF dont on se servira ensuite pour annoter les acteurs, les activités et les ressources en fonction des connaissances de l'ontologie. Les compétences sont créées en sélectionnant des connaissances de l'ontologie et en y ajoutant un niveau d'habileté et de performance.



**Figure 4** Modèle d'un scénario de conception d'un processus de gestion des connaissances.

- La troisième tâche consiste à *décrire les connaissances (et les compétences) des acteurs, activités et ressources* en utilisant un « annotateur sémantique ». C'est une responsabilité de l'équipe de conception, notamment d'un professionnel de la gestion documentaire. On peut aussi choisir d'annoter les ressources par les connaissances et les compétences qu'elles détiennent (dans le cas des acteurs) ou qu'elles requièrent ou visent dans le cas des activités et des ressources (compétences préalables ou visées). Une fois annotées, les ressources seront intégrées dans un référentiel qui permet de les retrouver par leurs annotations sémantiques. Ce sont ces annotations qui lient les composantes d'un scénario de travail et permettent d'offrir un soutien proactif aux acteurs.
- La quatrième tâche de l'équipe de conception consiste à *rechercher les ressources (y compris les acteurs) par leurs connaissances (ou leurs compétences)* pour les associer comme intrants aux activités prévues dans le scénario semi-formel de l'étape 1. Pour chaque activité, la sélection des ressources dépendra des connaissances (ou des compétences) préalables ou visées par l'activité et de celles dont disposent les acteurs qui réalisent l'activité. On sélectionne les ressources intrants ou à produire en naviguant dans les annotations sémantiques ou à l'aide d'un outil de recherche sémantique par connaissances ou par compétences.
- La cinquième tâche consiste à finaliser le scénario semi-formel à l'aide de l'éditeur de scénarios, en en y intégrant les ressources sélectionnées à l'étape précédente et en modifiant au besoin les activités et leurs acteurs responsables pour produire un scénario exécutable, équilibré en termes de connaissances et de compétences.

Ce processus-cadre de gestion des connaissances a été utilisé en partie dans un projet de gestion des connaissances dans une entreprise d'électricité. Une partie de l'ontologie résultant de la modélisation des connaissances contient des classes d'équipements, de diagrammes et de documentation nécessaires à la réalisation des tâches d'un technicien en automatismes de production d'électricité. Ce modèle a servi à référencer les documents contenus dans diverses bases de connaissances de l'organisation par les classes de cette ontologie. Cela a permis par la suite de les interroger comme s'il s'agissait d'une seule banque de documents.



**Figure 5** Une partie d'un scénario de travail d'un technicien d'un traducteur de voltage.

Des interrogations à partir des annotations sémantiques permettent de sélectionner et d'associer des collections de ressources aux tâches d'un technicien. Ces collections sont intégrées dans le processus de travail (figure 5). Par exemple, la collection de ressources N6 regroupe les ressources nécessaires à la « préparation du montage » pour les tests. De même, la collection de ressources N7 regroupe les intrants de la tâche « réaliser le test » et la collection N8 regroupe les intrants de la tâche « Ajuster le gain.... ». Selon l'interprétation du résultat du test, le technicien doit décider (choix exclusif) de lancer la compilation de la fiche vue de l'archivage ou de modifier les réglages pour ensuite refaire le test et les autres étapes subséquentes du processus.

Ce processus de travail fournit aux acteurs un environnement complet contenant l'ordonnancement des tâches, les collections de ressources qui peuvent être consultées pour les réaliser ou que les acteurs doivent produire.

## 4. Web social et communautés de pratique

On peut traiter la gestion des connaissances d'un point de vue individuel ou organisationnel, et aussi d'un point de vue technologique ou relatif aux ressources humaines. De ce dernier point de vue, sur un plan individuel, on s'intéressera à la formation alors que sur un plan organisationnel, le centre d'intérêt se déplace vers la communication, les interactions et les communautés de pratique.

Selon Prax, « la communauté professionnelle est l'essence même de l'organisation basée sur les flux de connaissances. Elle est très informelle et souvent spontanée, donc très difficile à manager<sup>9</sup> ». Selon Étienne Wenger, « si une entreprise ne reconnaît pas officiellement les communautés de pratique, elles existeront quand même. Mais elles ne produiront pas les bénéfiques que l'entreprise aurait pu en tirer<sup>10</sup> ».

Le praticien qui participe à une communauté apprend par les échanges avec ses collègues, résout des problèmes, identifie les expériences et les pratiques des autres. Il peut mieux ainsi appliquer les connaissances et les compétences acquises dans son contexte opérationnel et ses processus de travail. Que la communauté soit locale ou distribuée dans plusieurs édifices, plusieurs villes ou plusieurs pays ou continents, elle a besoin d'un environnement technologique riche pour soutenir sa fondation, son fonctionnement, son déploiement et sa productivité.

Dans cette section, nous nous intéressons à l'intégration des technologies sémantiques et du web social comme moyen de soutenir les communautés de pratique engagées dans les processus de gestion des connaissances.

### 4.1 Le web social

Le web social ou web 2.0 nous intéresse particulièrement en ce qu'il facilite la collaboration et le partage d'information, et donc le fonctionnement des communautés de pratique. Le web social est en pleine émergence. On ne compte plus les applications et les outils qui s'installent dans notre quotidien dont les plus connus sont Wikipédia, Facebook, Twitter, YouTube Flickr, delicious et les outils de blogue. En fait, tout comme le web sémantique, le web social se construit par-dessus le web initial et caractérise l'état actuel de l'ingénierie du web. Par ailleurs, le phénomène des *mashup* (mixage) fait en sorte que certains services offerts par différents sites peuvent être regroupés pour combiner les données de manière nouvelles et innovatrices.

Le web social a des applications sur les intranets tout comme sur l'internet. Dans les compagnies et les organisations, les applications qualifiées d'entreprise 2.0 regroupent des outils web 2.0 pour créer des environnements utilisés pour la gestion des connaissances, le travail collaboratif et le soutien aux communautés de pratique. Alors que sur le web, les

---

<sup>9</sup> Prax, J. Y. (2007). *Le Manuel du Knowledge Management – mettre en réseau les hommes et les savoirs pour créer de la valeur* (2<sup>e</sup> éd.). Paris : Dunod, p. 387.

<sup>10</sup> Wenger, E., McDermott, R. et Snyder, W. M. (2002). *Cultivating communities of practices* (1<sup>ère</sup> éd.), Harvard Business Review Press.

outils du web social sont orientés principalement vers le loisir, leurs usages sérieux sont en pleine émergence sur les lieux de travail ou dans les communautés scientifiques.

L'utilisation du web social pose toutefois certaines difficultés. Une limite actuelle des sites de réseautage social est leur isolement les uns des autres. Ce sont généralement des mondes fermés et des îlots de données indépendants contenant des connaissances complémentaires qui ne sont pas immédiatement accessibles d'un endroit ou l'autre. Par exemple, quelqu'un qui s'intéresse à un produit particulier, disons une caméra, pourrait trouver des commentaires sur Twitter, des photos sur Flickr et des vidéos sur YouTube, mais pourrait difficilement obtenir ces informations à partir de son propre réseau social. Plus le nombre de sites sociaux, de communautés et de services s'ajoutent au web social, plus leur absence d'interopérabilité devient évidente et problématique, rendant difficile la synergie des informations.

Une des raisons du cloisonnement des réseaux sociaux est le désir de leurs opérateurs de garder leurs utilisateurs captifs au sein du réseau. Un utilisateur ne peut pas bouger d'un réseau social à un autre sans perdre toute l'information, les liens et l'historique qu'il y a accumulés. Les technologies sémantiques peuvent être utilisées pour jeter des points entre les sources de données sociales, de sorte que plus les utilisateurs interagissent entre eux, plus les connaissances deviennent interreliées. Cette augmentation de la valeur intrinsèque d'un réseau ne se reflète pas toutefois dans la valeur de la compagnie qui gère le réseau, en regard de ses compétiteurs. Malgré tout, Facebook, par exemple, rend ses pages plus accessibles et également ses données au moyen de l'Open Graph Protocol<sup>11</sup>, sans doute parce que cela augmente le potentiel des ventes publicitaires.

Une raison plus technique du manque d'interopérabilité tient à l'absence de standards pour l'échange d'informations et de connaissances entre la plupart des applications du web social. Le RSS (Really Simple Syndication ou RDF Site Syndication) est le premier pas dans la bonne direction, fournissant un format pour annoncer la mise à jour de contenu tels les messages de blogues. Il demeure toutefois difficile d'identifier et de réutiliser des données communes entre les blogues, les forums, les wikis et les sites de réseautage social. À titre d'exemple, un utilisateur peut créer des contenus sur plusieurs de ces outils, mais un autre utilisateur ne peut repérer sa contribution sur l'ensemble de ces outils.

## 4.2 Intégration du web social et du web sémantique

Le web sémantique fournit des outils nécessaires pour résoudre ces difficultés du web social en définissant des standards pour l'échange d'informations et l'interopérabilité entre les applications et les sites web. Un certain nombre de vocabulaires du web sémantique sont déjà disponibles à cette fin :

- RSS 1.0<sup>12</sup> pour la syndication des informations de sources variées repose sur un modèle RDF, contrairement à d'autres versions du RSS;
- FOAF<sup>13</sup> (Friend of a Friend) permet d'exprimer des profils personnels et de l'information provenant des réseaux sociaux;

---

<sup>11</sup> Site web : <http://www.opengraphprotocol.org/>

<sup>12</sup> Site web : <http://web.resource.org/rss/1.0/>

- SIOC<sup>14</sup> (Semantically Interlinked Online Communities) sert à relier les communautés et à distribuer les conversations sur le web.

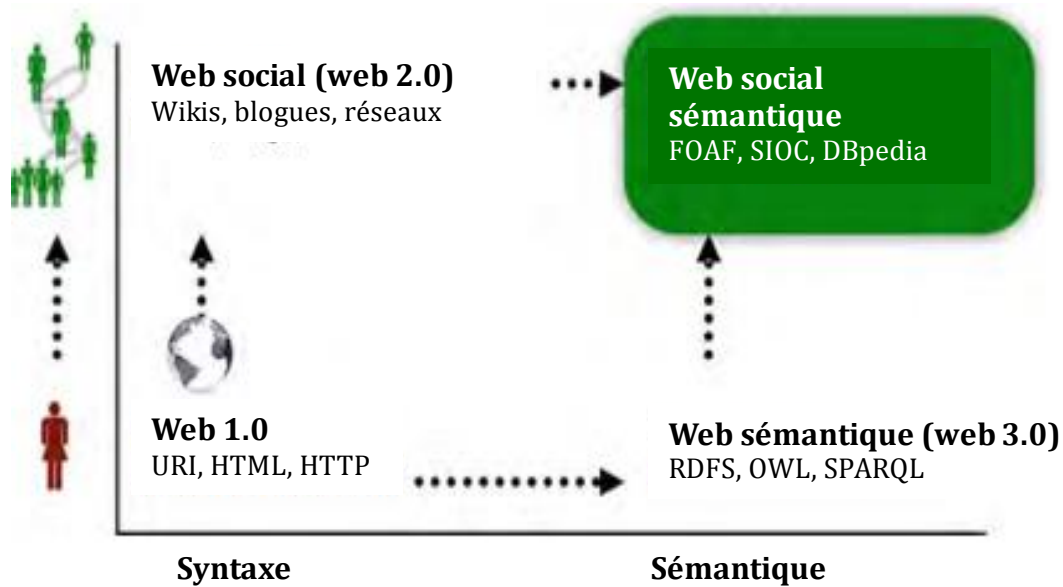


Figure 6 L'évolution vers le web social sémantique<sup>15</sup>.

L'application des technologies sémantiques au web social nous conduit au web social sémantique. Comme il est illustré sur la figure 6, le web social ou web 2.0 est une évolution du web initial (web 1.0) qui valorise les usages collaboratifs plutôt qu'individuels. Le web sémantique ou web 3.0 est une évolution orthogonale d'une tout autre nature, mettant l'accent sur les connaissances et la sémantique plutôt que sur la syntaxe des documents web. Le web social sémantique est qualifié par certains de web 4.0 parce qu'il combine les deux types d'évolution, créant un réseau de connaissances riches et interreliées, intégrant les applications et l'aspect social et collaboratif du web 2.0 avec les formats et les langages de représentation des connaissances du web 3.0.

Cette évolution est d'une grande importance pour la gestion des connaissances, compte tenu de la part grandissante que prennent le travail collaboratif et les communautés de pratique dans les organisations. Les technologies sémantiques peuvent soutenir les communautés en ligne en liant les documents, les données et les applications qui résultent d'une variété d'interactions décrites dans des formats utilisables par les ordinateurs. Inversement, les communautés en ligne apportent au web sémantique des données

<sup>13</sup> Brickley, D. et Miller, L. (2004). FOAF vocabulary specification. Namespace document, FOAF project. Récupéré le 20 octobre 2013 de : <http://xmlns.com/foaf/0.1/>

<sup>14</sup> Site web : <http://www.sioc-project.org>

<sup>15</sup> Traduction d'un figure de Domingue, Fensel et Hencler (dir.). (2001). *Handbook of semantic web technologies*. Springer, p. 472.

provenant des utilisateurs qui travaillent en réseau. Ainsi, l'intégration entre le web social et le web sémantique est doublement intéressante :

- D'une part, des efforts sont entrepris pour utiliser les technologies sémantiques dans la modélisation des données sociales. Des ontologies comme FOAF et SIOC permettent de représenter les données sociales dans des vocabulaires communs et partagés, favorisant l'interopérabilité et la portabilité entre les applications.
- D'autre part, en analysant les informations exprimées par le grand nombre d'utilisateurs des services du web 2.0, on peut créer une masse de données pour le web sémantique et rendre les applications plus « intelligentes » que par le passé.

Les consensus dans les communautés d'utilisateurs créent du sens et des connaissances partagées autour desquelles ces utilisateurs resserrent leurs liens de réseautage social. En même temps, ils fournissent la masse de données nécessaire pour amorcer les applications du web sémantique. En retour, celles-ci ouvrent alors des possibilités nouvelles impossibles à réaliser uniquement avec les services actuels du web social.

Nous allons maintenant explorer des ponts entre les deux technologies, soit les ontologies pour le web social et les applications du web social qui intègrent des technologies sémantiques.

#### **4.3 Ontologies pour le web social : FOAF, SIOC et folksonomies**

L'utilisation de langages standards de représentation des connaissances, tels que RDFS ou OWL, permet de partager les ontologies entre les services informatiques, rendant les données interopérables entre les applications distribuées sur le web. Dans le cadre du web social sémantique, les ontologies peuvent servir à représenter de façon uniforme les différentes entités produites et partagées dans les sites sociaux, soit les groupes, les personnes, les documents, les messages, les annotations.

**FOAF** est l'un de ces vocabulaires largement utilisés pour décrire les personnes et les relations entre elles, ainsi que les objets qu'elles créent et leurs centres d'intérêt. Le tableau suivant décrit les principales classes et propriétés du vocabulaire FOAF.

Chacun peut ainsi créer son propre fichier FOAF, se décrivant lui-même ainsi que son réseau social. L'information de plusieurs fichiers FOAF peut être facilement combinée pour obtenir une vue plus complète du réseau, sans l'aide d'une base de données centralisée, en appliquant les principes du web de données liées. FOAF peut aussi être intégré avec tout autre vocabulaire du web sémantique tels SIOC, SKOS, DBpedia, GeoNames, etc.

**Tableau 1** Principales classes et propriétés du vocabulaire FOAF

CLASSES	foaf:Person	Pour décrire les personnes.
	foaf:OnlineAccount	Pour énumérer les comptes d'une personne sur le web.
	foaf:Document	Pour décrire les documents que créent les personnes.
PROPRIÉTÉS	foaf:knows	Pour créer un lien entre les personnes, formant ainsi un réseau.
	foaf:mbox_sha1sum	Pour identifier quelqu'un par une adresse courriel.
	foaf:topic_interest	Pour déterminer les ressources et les concepts qui intéressent une personne, notamment les sujets de DBpedia.

**SIOC** a pour objectif de relier des contenus de communautés en ligne tels que les blogues, la messagerie et les sites sociaux, en décrivant la structure et les activités de communautés en ligne. En combinant ce vocabulaire avec FOAF et SKOS pour l'organisation des connaissances, SIOC permet aux développeurs de lier des éléments de contenu créés dans des sites de messagerie, des forums, des blogues avec d'autres éléments, des personnes décrites par leur compte utilisateur et des sujets décrits par des taxonomies ou des concepts représentés par leurs URI.

L'ontologie SIOC a été adoptée dans le cadre d'environ 60 applications ou modules, de certains outils d'exportation pour des plateformes du web social jusqu'à des applications en recherche neuromédicale. À titre d'exemple, Yahoo! a utilisé SIOC dans son application SearchMonkey<sup>16</sup> comme vocabulaire de référence pour décrire les activités de communautés en ligne, en combinaison avec FOAF pour décrire les participants. Un autre aspect intéressant de SIOC est son utilisation pour modéliser les interactions sociales dans des communautés de pratique dans les intranets des organisations ou pour soutenir les discussions dans des communautés scientifiques.

Le système SIOC se compose de l'ontologie noyau SIOC, d'un schéma RDF qui comprend 11 classes et 53 propriétés, ainsi que de cinq modules périphériques : SIOC Access, SIOC Argumentation, SIOC Services, SIOC Types, et SWAN/SIOC. Les principaux termes de l'ontologie noyau sont présentés sur le graphe RDF-S de la figure 7, construit à l'aide du logiciel G-MOT<sup>17</sup>.

---

<sup>16</sup> Site web : <http://developer.yahoo.com/searchmonkey/>

<sup>17</sup> Les classes de l'ontologie y sont représentées par des rectangles et les propriétés par des hexagones, liées à leurs classes domaine et codomaine par les liens R. Le lien S représente la relation entre les sous-classes et les classes.

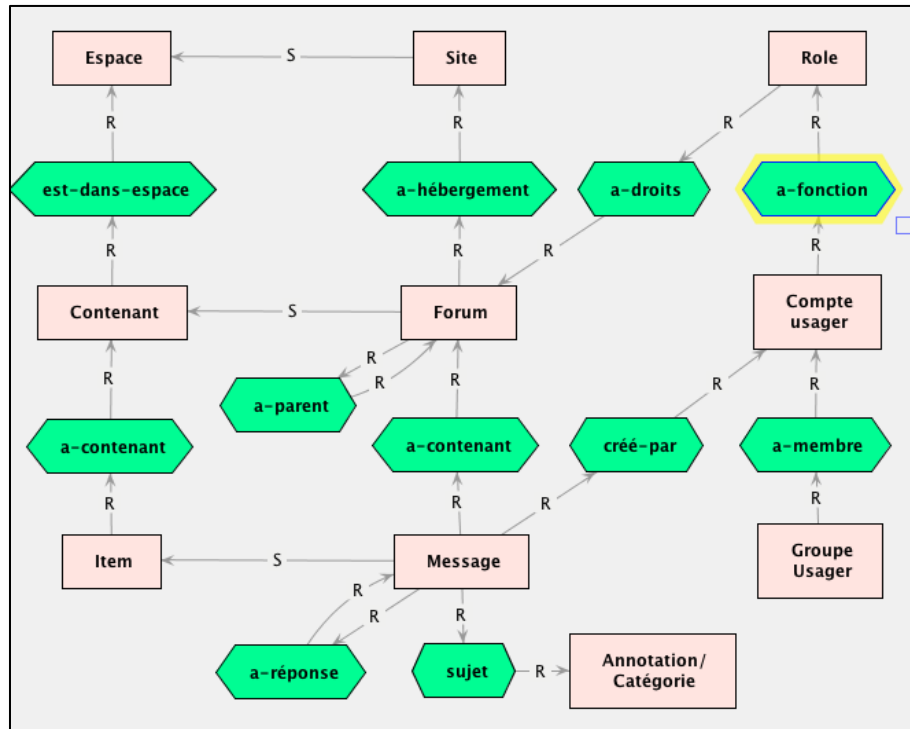


Figure 7 L'ontologie SIOC (traduction française des termes anglais).

Comme nous pouvons le constater, les termes ont été choisis de façon générique de façon à s'appliquer à un grand nombre de situations d'échanges de contenu. Le modèle de base de SIOC repose sur des messages (*post*) créés par des comptes utilisateurs identifiant des personnes. Ces personnes sont membres de groupes d'utilisateurs et ont des rôles dans un forum qui peut être parent d'autres forums. Les forums sont hébergés sur des sites et contiennent des messages qui peuvent avoir des réponses qui, elles-mêmes, peuvent entraîner d'autres réponses.

Les propriétés définies dans SIOC permettent de décrire des relations entre ces éléments de contenu et leurs attributs. Par exemple, la propriété `sioc:has_reply`, relie les messages dont l'un est une réplique de l'autre; la propriété `sioc:has_creator` et `foaf:maker` identifie les auteurs et fournit de l'information additionnelle à leur sujet; la propriété `sioc:topic` pointe vers une ressource décrivant le sujet d'un élément de contenu.

En parallèle, avec l'arrivée de nouveaux sites sociaux différents des forums ou des listes de courriels, ces concepts sont devenus des sous-classes de concepts de plus haut niveau : les espaces de données, les contenants et les items de contenu. Ces classes plus larges permettent de structurer l'information dans les sites de communautés en ligne et d'établir des distinctions entre les différentes catégories d'éléments de contenu, au-delà des messages.

Un espace de données (`sioc:space`) est un lieu où réside des données, par exemple un site web, un ordinateur personnel, un espace sur un serveur contenant des fichiers partagés, etc. Un espace peut loger un ensemble de contenants regroupant des items, par exemple des

messages, mais aussi d'autres éléments de contenu. Des sous-classes de contenants autres que des forums peuvent être définis pour regrouper d'autres types d'éléments que des messages. Par exemple, un carnet d'adresses qui décrit une collection de contacts professionnels (`sioc:item`) est un type de contenant (`sioc:Container`) qui est différent d'un forum de discussion.

Un module du système SIOC, SIOC Types, offre plus de sous-classes des concepts de base du noyau SIOC, introduisant différentes catégories d'objets sociaux. En voici quelques exemples :

- `AddressBook`, `AnnotationSet`, `AudioChannel`, `BookmarkFolder`, `Briefcase`, `EventCalendar`, etc. sont des sous-classes de `sioc:Container`.
- `ChatChannel`, `MailingList`, `MessageBoard`, `Weblog` sont des sous-classes de `sioc:Forum`.
- `BlogPost`, `BoardPost`, `Comment`, `InstantMessage`, `MailMessage`, `WikiArticle` sont des sous-classes de `sioc:Post`.
- D'autres termes comme `Answer`, `BestAnswer`, `Question` se sont également ajoutés pour traiter des sites de question-réponse sur le web sémantique.

Un autre module, le SIOC Services, assure le lien avec les interfaces à des services web, utilisant les protocoles SOAP ou REST pour le traitement des éléments de contenu. Ce module permet d'indiquer qu'un service web est associé ou fait partie d'un `sioc:Site`, à ne pas confondre avec la définition du service web. Une propriété `sioc_s:service_definition` sert à relier un `sioc_s:Service` à sa définition, par exemple en WSDL (Web Service Definition Language).

Le SIOC Access Module sert à définir l'information à propos des droits d'accès des utilisateurs en fonction de leurs rôles et du statut des éléments de contenu. Ces derniers peuvent être publics, privés, brouillons. Le module SIOC Argument permet de formuler des accords et des désaccords entre des éléments de contenu. Il est fondé sur des modèles d'argumentation.

**Folksonomies.** Un *marqueur* (ou une étiquette, un libellé ; *tag* en anglais) est un mot clé (signifiant) ou un terme associé ou assigné à de l'information (par exemple une image, un article, ou un clip vidéo), qui décrit ainsi l'objet et permet une classification des informations basées sur les mots clés. Bien que certains opposent la modélisation des données par ontologie et les folksonomies qui résultent de l'association libre de marqueurs ou de mots clés, les deux approches peuvent être combinées de façon utile.

Une méthode pour ce faire consiste à utiliser les principes de la modélisation des langages RDFS ou OWL pour représenter les marqueurs, les actions de marquage et les nuages de marqueurs. La recherche à base de marqueurs est actuellement la seule façon de trouver les contenus marqués. Comme nous l'avons indiqué précédemment, cette pratique peut poser des problèmes de polysémie, si des personnes utilisent le même terme avec des significations différentes, ou encore des problèmes de synonymie, lorsqu'elles utilisent différents termes pour la même signification.

En publiant les marqueurs et le contenu marqué par des triplets RDF (marqueur, référence, item-contenu), on peut établir des liens avec d'autres données sémantiques et les réutiliser d'une application à l'autre. On peut alors utiliser les technologies sémantiques pour répondre à des questions comme : « Quels items sont référencés par "Apple" sur toute plateforme? », « Quels sont les derniers dix marqueurs utilisés par Jean sur la plateforme delicious? », « Affichez la listes des marqueurs utilisés à la fois par Alex sur SlideShare et par Jean sur Flickr. »

Gruber<sup>18</sup> a proposé l'une des premières approches pour modéliser les folksonomies en utilisant une ontologie. En plus d'intégrer les marqueurs dans des triplets les liant aux objets référencés, Gruber a proposé d'étendre ce modèle par (a) un attribut « espace » qui représente le site web où le *tagging* s'est produit et (b) une valeur de « polarité » pour tenir compte du phénomène du *spam*. Sa proposition fournit un modèle complet pour représenter les actions de *tagging*, tout en tenant compte du fait que plusieurs marqueurs différents peuvent se référer au même concept. Ceci montre la nécessité de déterminer une sémantique commune entre ces marqueurs.

La TAG Ontology<sup>19</sup> a été le premier modèle RDF pour représenter les actions de marquage. Il est fondé sur les idées initiales de Gruber. Cette ontologie définit des classes `Tag` et `Tagging` avec des propriétés de ces classes. Pour représenter les utilisateurs réalisant des *tagging* action, l'ontologie fait appel au vocabulaire FOAF. Le fait de définir une classe `Tag` de marqueurs implique que ceux-ci auront tous une URI, ce qui les rend apte à servir à la fois comme sujet et comme objet de triplets RDF. Puisque la classe `Tag` est définie comme une sous-classe de la classe `skos:Concept`, les marqueurs peuvent être liés ensemble, par exemple pour dire que « géométrie » est plus spécifique que « mathématique ».

#### 4.4 Applications du web social sémantique

Nous allons maintenant présenter des applications qui intègrent des technologies sémantiques à des outils de web social. On retrouve dans cette catégorie des outils de blogues sémantiques, de microblogues sémantiques, de wikis sémantiques, des outils de signets sémantiques dans les réseaux sociaux, des sites de revues et des sites de partage scientifique.

##### Blogues sémantiques

Un blogue est essentiellement un site web qui affiche un journal dont les entrées sont disposées en ordre chronologique inverse. Les entrées contiennent des textes, des liens à d'autres sites, et parfois des images et des vidéos. Les blogues collectent et créent de grandes quantités de données sur un ordinateur, mais ces données sont souvent enfermées dans des applications propriétaires ou sur le disque dur d'un appareil personnel, dans des carnets d'adresses personnels, un agenda ou des fichiers audio ou vidéo.

---

<sup>18</sup> Gruber, T. (2007). Ontology of folksonomy : A mash-up of apples and oranges. *International Journal on Semantic Web & Information Systems*, 3(2), 1-11.

<sup>19</sup> Site web : <http://www.holygoat.co.uk/projects/tags/>

Les blogues sémantiques visent à publier ces données sur le web de données liées. Par exemple, dans un blogue sémantique, un utilisateur écrit un message et l'annote par des métadonnées prises dans son carnet d'adresses. Il publie le message sur le web et quelqu'un d'autre lisant le message peut réutiliser les métadonnées dans ses propres applications, utilisant le web comme un *clipboard*.

SparqlPress<sup>20</sup> est un prototype qui opère ce type de fonction. Ce n'est pas un outil de blogues, mais un *plug-in* en code libre pour la plateforme de blogues WordPress qui permet de produire, intégrer et réutiliser des données RDF. SparqlPress utilise les vocabulaires FOAF, SIOC et SKOS. Une fonctionnalité intéressante est l'utilisation de FOAF et des identités openID pour afficher de l'information additionnelle à propos d'un utilisateur du blogue ou pour bloquer d'éventuels *spam* ou des messages d'utilisateurs indésirables.

### Microblogues sémantiques

Les outils de microblogging permettent de publier des messages très courts, axés sur l'échange d'information rapide en temps réel. Ces messages sont aussi appelés *tweet* sous l'influence de Twitter, le plus populaire de ces outils. Cette méthode de communication est aussi prometteuse dans les environnements organisationnels, facilitant la communication informelle, la formation et la propagation rapide de l'information. Pour une organisation, le microblogging peut favoriser la propagation rapide des questions-réponses et les notifications de rencontres ou d'horaires des livrables. IBM a récemment déployé à l'interne un système de microblogging appelé « Blue Twit ». On peut penser qu'une tendance vers ce type d'outil émergera dans les années qui viennent, de façon similaire à l'essor des blogues, des wikis et d'autres services de type entreprise 2.0.

La plateforme de microblogging SMOB<sup>21</sup> est un exemple d'un outil ouvert pour une plateforme décentralisée de publication distribuée de *tweets*, contrairement à Twitter qui demeure un outil propriétaire centralisé. SMOB démontre qu'il est possible de fournir à des usagers des outils pour contrôler, partager, remixer leurs propres données, sans être dépendant d'un service intermédiaire. Le contenu publié avec SMOB demeure la propriété de l'utilisateur qui l'a créé, localement, sur son propre appareil.

Cette plateforme fait elle aussi appel aux vocabulaires FOAF et SIOC pour modéliser les microblogueurs, leurs propriétés, leurs informations de compte et de services. Une multitude de services de publication peuvent capter les serveurs agrégateurs choisis par chaque utilisateur. La vue agrégée des messages des microblogues offre aux utilisateurs une interface conviviale pour afficher les données RDF complexes provenant de sources distribuées. De plus, les messages d'un microblogue peuvent inclure des marqueurs géographiques, lesquels combinés aux données de GeoNames permettent l'affichage visuel sur une carte géographique.

---

<sup>20</sup> Site web : <http://bzi.mfd-consult.dk/sparqlpress>

<sup>21</sup> Site web : <http://smob.sioc-project.org>

## Wikis sémantiques

Il y a maintenant des centaines de logiciels de type wiki dont le plus connu est MediaWiki, le logiciel qui maintient Wikipedia, l'encyclopédie du web. Les wikis sont des sites collaboratifs servant à une variété d'usages : créer des dictionnaires ou des glossaires, référencer des bibliographies, organiser des événements, préparer des soumissions de projets, soutenir des équipes de développement de logiciels. Un wiki est créé par une communauté dont les membres éditent des pages interreliées par des hyperliens. Tout membre de la communauté peut éditer un article d'un wiki ou en créer de nouveaux. Si une erreur est introduite par quelqu'un, le wiki contient un historique des modifications de façon à ce que le contenu puisse être rétabli. Le contenu d'un wiki est autorégulé par l'émergence d'un consensus dans la communauté établie, notamment au moyen de pages de discussions.

Contrairement à ce qui se produit dans les blogues sémantiques, des wikis tels que Wikipedia contiennent depuis leurs débuts des métadonnées structurées sous la forme de gabarits, lesquelles ont servi d'ailleurs à automatiser la création du vocabulaire DBpedia qui reprend, sous formes de triplets RDF, une partie substantielle du contenu de Wikipedia.

Plus récemment, une vingtaine de wikis sémantiques<sup>22</sup> sont apparus pour répondre au besoin d'accroître la structuration des wikis. Les wikis sémantiques offrent aux utilisateurs la création d'annotations sémantiques partout dans le texte d'un article pour offrir des accès internes plus ciblés et une réutilisation externe de l'information. Ces annotations permettent en général de créer des instances d'ontologies de domaines établies a priori ou qui émergent de l'usage du wiki (des folksonomies). Que les annotations soient internes à une page ou servent à fournir des métadonnées de la page, le système peut afficher ou permettre la navigation entre les pages à partir des liens RDF ou en intégrant des requêtes dans les pages.

Un wiki sémantique doit offrir un modèle sous-jacent des connaissances décrites dans ses pages pour identifier plus d'information (de métadonnées) à propos des pages et de leurs interrelations. Ce modèle de connaissances doit être exprimé dans un langage formel tel que RDFS ou OWL, de sorte que les programmes puissent, au moins partiellement, effectuer des traitements et faire des déductions.

Par exemple, un wiki sémantique doit savoir que lorsqu'une page parle des « pommes », il s'agit bien d'un fruit et non d'ordinateurs, de façon à afficher d'autres pages relatives aux pommes ou aux fruits. De plus, un wiki sémantique doit pouvoir répondre à des questions comme : « Quelles sont les pages dont le créateur est Jean Trudel? » ou « Quelles sont mes tâches à faire dans le projet? »

Il existe déjà plusieurs wikis sémantiques tels OntoWiki<sup>23</sup>, SweetWiki<sup>24</sup> et Semantic MediaWiki<sup>25</sup>, le plus largement utilisé. Il s'agit d'une extension de MediaWiki, le système à

---

<sup>22</sup> Site web : [http://semanticweb.org/index.php/Semantic\\_Wiki\\_State\\_Of\\_The\\_Art](http://semanticweb.org/index.php/Semantic_Wiki_State_Of_The_Art)

<sup>23</sup> Site web : <http://www.ontowiki.net>

<sup>24</sup> Site web : <http://sweetwiki.inria.fr/ontology>

<sup>25</sup> Krötzsch, M., Vrandečić, D. et Völkel, M. (2006). Semantic MediaWiki. *Proceedings of the Fifth International Semantic Web Conference (ISWC 2006)*, Athens, GA. Lecture Notes in Computer Science, vol. 4273, p. 935–942. Springer, Heidelberg.

la base de Wikipedia. Semantic MediaWiki permet l'intégration de données sémantiques décrivant les liens d'une page à une autre et les attributs d'une page du wiki. Le logiciel est complètement ouvert quant aux termes utilisés pour annoter le contenu, de sorte que le modèle de données et les ontologies sous-jacents évoluent selon les choix des utilisateurs.

Prenons un exemple. Dans Wikipedia, il y a une page à propos de l'auteur J. K. Rowling. Cette page possède un lien vers les livres qu'elle a écrits, un autre lien vers Edinbourg, son lieu de résidence, et un lien vers Scholastic Press, son éditeur. Dans un wiki traditionnel, on ne peut faire des recherches telles que : « Quels sont les livres écrits par J. K. Rowling? », « Quels sont les auteurs qui vivent en Écosse? » ou « Quels sont les autres auteurs publiés chez Scholastic Press? ». Dans le Semantic MediaWiki, on peut obtenir des réponses à ses questions en insérant dans la page Wikipedia de J. K. Rowling des liens tels que `[[author of::Harry Potter and the Deathly Hallows]]` ou `[[birthdate::1965-07-31]]`. De tels attributs permettront de répondre à la question « Quels sont les auteurs qui ont plus de 40 ans? » et d'obtenir une liste d'auteurs dont J. K. Rowling.

On passe alors d'une approche par documents et hyperliens dans un wiki traditionnel à une approche par concepts et relations dans un wiki sémantique. Les données dans le wiki sont exportées en utilisant le modèle de données RDF et les vocabulaires sous format RDFS ou OWL utilisés dans les annotations. Le Semantic MediaWiki est soutenu par une communauté active de développeurs qui créent des extensions au système de base, lequel est déjà utilisé dans une centaine de sites. Ces extensions améliorent l'édition des wikis, les recherches graphiques ou par mots clés, l'intégration avec des services de visualisation sur des cartes géographiques. Le système peut être localisé en plus de 40 langages.

### Signets sémantiques dans les réseaux sociaux

Des sites de réseautage social comme Facebook ou de réseautage professionnel comme LinkedIn servent à partager des informations sur soi avec des « amis » ou des relations professionnelles, par l'envoi de message ou de documents et divers outils de partage de connaissances. Il existe un outil de réseautage sémantique sur le web, semblable à Facebook ou LinkedIn, mais qui fait appel aux technologies sémantiques. Twine<sup>26</sup> applique des techniques de traitement du langage naturel pour assurer un marquage sémantique des textes à l'aide d'une ontologie regroupant quelques milliers de concepts. Le site de réseautage social est lui-même généré par une ontologie, ainsi que le modèle de données qui établit des liens entre les personnes et les sujets qui peuvent les intéresser. Un outil d'édition d'ontologies simples est offert aux utilisateurs dans le but de fournir une infrastructure extensible.

Un outil de recherche sémantique est aussi disponible. Par exemple, les signets peuvent être filtrés selon les organisations concernées ou les utilisateurs selon leurs lieux de naissances. Les données sémantiques de Twine peuvent être obtenues en format RDF pour être réutilisées ailleurs.

---

<sup>26</sup> Site web: <http://www.twine.com>

## Sites de revues

Revyu.com<sup>27</sup> est un service en ligne qui sert à créer des revues de différentes entités, par exemple, des articles de conférences ou des restaurants. Il utilise des principes communs aux applications du web social tels que des marqueurs (*tag*), des cotes d'évaluation et un outil JavaScript qui permettent de créer des signets en naviguant sur le web, lesquels serviront à construire de nouvelles revues.

Le site repose entièrement sur un modèle de données RDF. Chaque revue est modélisée en utilisant le vocabulaire Review. Les marqueurs sont représentés en utilisant l'ontologie Tag. Revyu.com fournit un point d'accès SPARQL pour la recherche dans ses données et permet la réutilisation des marqueurs de données dans tout autre application, ainsi que la création de *mashup* avec des contenus existants et la mise en relation avec d'autres ensembles de données tels que ceux du web de données liées. Par exemple, les revues de films sont liées à des URI de DBpedia. L'utilisation de profils FOAF pour représenter les utilisateurs évite de remplir les informations personnelles demandées dans beaucoup de plateformes.

## Conclusion

L'ensemble des applications présentées ici démontre que les applications du web sémantique ont dépassé le stade du laboratoire. En lien ou non avec le web social, les applications des technologies sémantiques se développent extrêmement rapidement et rendent déjà d'immenses services pour faire du web un outil plus convivial et plus efficace.

---

<sup>27</sup> Heath, T. et Motta, E. (2007). Revyu.com : A reviewing and rating site for the web of data. In: *Proceedings of the Sixth International Semantic Web Conference and Second Asian Semantic Web Conference (ISWC/ASWC 2007)*, Busan. Lecture Notes in Computer Science, vol. 4825, pp. 895–902. Springer, Heidelberg.